6.896 Sublinear Time Algorithms	October 7, 2004
Lecture 8	
Lecturer: Ronitt Rubinfeld	*Scribe: Kayi Lee*

1 Review

In the previous lecture, we introduced the Szemerédi's Regularity Lemma, which states that all graphs can be approximated by random-looking graphs in a certain sense.

Definition 1 *A pair of disjoint vertex sets (A, B) is γ-regular if, for all $A' \subseteq A, B' \subseteq B$ such that $|A'| \geq \gamma|A|$ and $|B'| \geq \gamma|B|$, we have $|d(A, B) - d(A', B')| < \gamma$, where $d(A, B) = \frac{e(A,B)}{|A||B|}$ is the density between the sets A and B.*

Note that if (A, B) is γ-regular, (A, B) is also γ'-regular for all $\gamma \leq \gamma' \leq 1$ by definition.

Lemma 2 (Szemerédi's Regularity Lemma) *For all $m, \epsilon > 0$, there exists T such that if $G = (V, E)$ with $|V| > T$ and $\mathcal{A}$ is a equipartition of V into m sets, then there exists a equipartition $\mathcal{B}$, a refinement of $\mathcal{A}$, with k sets such that:*

1. *$m \leq k \leq T$; and*
2. *there are fewer than $\epsilon\binom{k}{2}$ pairs of set partitions that are not ϵ-regular*

We also proved the following lemma, which states that a triplet of pairwise regular vertex sets behaves like a random graph with respect to the number of distinct triangles in the graph.

Lemma 3 (Komlos and Simonovits) *For all $\eta > 0$, there exists γ^Δ and δ^Δ such that if A, B, C are disjoint subsets of V and each pair is γ^Δ-regular with density at least η, then G contains at least $\delta^\Delta|A||B||C|$ distinct triangles with vertex from each of A, B and C.*

Note that the constants T, γ^Δ and δ^Δ in the above two lemmas depend on the choices of m, ϵ and η, but not the size of the graph. In the rest of the note, we will denote them as as $T(m, \epsilon), \gamma^\Delta(\eta)$ and $\delta^\Delta(\eta)$ to remind readers of the parameters these constants depend on.

2 A 1-sided Triangle-Free Property Tester

We have described a 1-sided triangle-free property tester that is based on the following theorem by Alon (see also [Alon Fischer Krivilevich Szegedy]):

Theorem 4 *For all ϵ, there exists δ such that any n-vertex graph $G = (V, E)$ that is ϵ-far from triangle-free in the adjacent matrix model has at least $\delta\binom{n}{3}$ triangles.*

Proof We start with an arbitrary equipartition $\mathcal{A}$ with $\frac{5}{\epsilon}$ sets of vertices. Pick $\epsilon' = \min\{\frac{\epsilon}{5}, \gamma^{\Delta}(\frac{\epsilon}{5})\}$ and $T = T(\frac{5}{\epsilon}, \epsilon')$, where $\gamma^{\Delta}(\frac{\epsilon}{5})$ and $T(\frac{5}{\epsilon}, \epsilon')$ are the constants given by Lemmas 3 and 2 respectively. By the Regularity Lemma, we can refine $\mathcal{A}$ to get another equipartition $\mathcal{B} = \{V_1, V_2, \ldots, V_k\}$ such that $\frac{5}{\epsilon} \le k \le T$ and fewer than $\epsilon'\binom{k}{2}$ pairs of sets are not ϵ'-regular. Also, note that the size of each set $|V_i|$ is $\frac{n}{k}$, where $\frac{n}{T} \le \frac{n}{k} \le \frac{\epsilon n}{5}$.

Now given the equipartition $\mathcal{B} = \{V_1, V_2, \ldots, V_k\}$, consider an edge $\{u, v\}$ where $u \in V_i$ and $v \in V_j$. We call the edge $\{u, v\}$ *useful* iff:

1. $V_i \neq V_j$; and
2. (V_i, V_j) is ϵ'-regular; and
3. the density between V_i and V_j is at least $\frac{\epsilon}{5}$.

An edge is *non-useful* if it is not useful. The following lemma says that there are fewer than $\epsilon\binom{n}{2}$ non-useful edges in the graph. We will prove this lemma later.

Lemma 5 *There are fewer than $\epsilon\binom{n}{2}$ non-useful edges with respect to the equipartition $\mathcal{B}$.*

Suppose a graph G is ϵ-far from triangle-free. We construct a new graph G' by removing all the non-useful edges from G. Since there are fewer than $\epsilon\binom{n}{2}$ non-useful edges by Lemma 5, G' is not triangle-free. In other words, there exist $u \in V_i, v \in V_j, w \in V_k$ where $\{u, v\}, \{v, w\}$ and $\{u, w\}$ are useful edges in G. By the definition of useful edges, we have:

1. V_i, V_j and V_k are distinct,
2. V_i, V_j and V_k are pairwise ϵ'-regular, which implies they are $\gamma^{\Delta}(\frac{\epsilon}{5})$-regular since $\epsilon' \le \gamma^{\Delta}(\frac{\epsilon}{5})$,
3. the density between each pair from V_i, V_j and V_k is at least $\frac{\epsilon}{5}$.

Therefore, by Lemma 3 it follows that the number of distinct triangles in G is at least

$$\begin{aligned} \delta^{\Delta}(\frac{\epsilon}{5})|V_i||V_j||V_k| &= \delta^{\Delta}(\frac{\epsilon}{5}) \cdot (\frac{n}{k})^3 \\ &\ge \delta^{\Delta}(\frac{\epsilon}{5})\frac{n^3}{T^3} \\ &= \delta' n^3, \qquad \text{where } \delta' = \frac{\delta^{\Delta}(\frac{\epsilon}{5})}{T^3}. \end{aligned}$$

■

Finally, we prove Lemma 5.

Proof of Lemma 5 If an edge $\{u, v\}$, where $u \in V_i$ and $v \in V_j$, is non-useful, it has to satisfy one of the three cases:

1. u and v belong to the same partition. Each node has at most $\frac{n}{k} - 1$ neighbors that belong to the same partition. Therefore the number of edges of this kind is at most
$$\begin{aligned} n \cdot (\frac{n}{k} - 1) &= \frac{n(n-k)}{k} \\ &\leq \frac{n(n-1)}{k} \\ &= \frac{2}{k}\binom{n}{2} \\ &\leq \frac{2\epsilon}{5}\binom{n}{2}. \end{aligned}$$

2. $V_i \neq V_j$ and (V_i, V_j) is not ϵ'-regular. By the Regularity Lemma there are at most $\epsilon'\binom{k}{2}$ such pairs. Since each pair contains at most $(\frac{n}{k})^2$ cross edges, the number of edges of this kind is at most
$$\begin{aligned} \epsilon'\binom{k}{2}(\frac{n}{k})^2 &\leq \epsilon'\binom{k}{2}(\frac{n}{k})(\frac{n-1}{k-1}) \\ &= \epsilon'\binom{n}{2} \\ &\leq \frac{\epsilon}{5}\binom{n}{2}. \end{aligned}$$
The last inequality is due to the fact that $\epsilon' = \min\{\frac{\epsilon}{5}, \gamma^{\Delta}(\frac{\epsilon}{5})\}$.

3. $V_i \neq V_j$ and the density between V_i and V_j is less than $\frac{\epsilon}{5}$. In this case, there are less than $(\frac{\epsilon}{5})(\frac{n}{k})^2$ cross edges between V_i and V_j. Since there are $\binom{k}{2}$ different pairs of V_i and V_j, the number of edges of this kind is at most
$$\begin{aligned} \binom{k}{2}(\frac{\epsilon}{5})(\frac{n}{k})^2 &\leq \binom{k}{2}(\frac{\epsilon}{5})(\frac{n}{k})(\frac{n-1}{k-1}) \\ &= \frac{\epsilon}{5}\binom{n}{2}. \end{aligned}$$

Therefore, the total number of non-useful edges is at most $\frac{2\epsilon}{5}\binom{n}{2} + \frac{\epsilon}{5}\binom{n}{2} + \frac{\epsilon}{5}\binom{n}{2} < \epsilon\binom{n}{2}$.

■

The property tester picks three distinct vertices at random, and check if they form a triangle. If the graph G is triangle free, the tester will never detect a triangle. So it will always pass. On the other hand, if the graph is ϵ-far from triangle-free, Theorem 4 implies that the probability of detecting a triangle is at least δ. Therefore, by repeating the experiment independently for $\Theta(\delta^{-1})$ times, the probability of detecting a triangle in G is at least $\frac{2}{3}$. This gives us a 1-sided triangle-free property tester.

3 A Characterization of Bipartiteness by Query Complexity

The following theorem is proved by Alon, which gives an interesting characterization of bipartiteness of a graph by the query complexity for testing a certain class of graph properties in the adjacency matrix model.

Theorem 6 *In adjacency matrix model, Let H be fixed with $|H| = h$ and P_H be the property that a graph does not contain H as a subgraph. Then the following holds:*

1. *If H is bipartite,*
   - *(a) there exists a 2-sided error tester for P_H with $O(\epsilon^{-1})$ queries; and*
   - *(b) there exists a 1-sided error tester for P_H with $O(h^2(\frac{1}{2\epsilon})^{\frac{h^2}{4}})$ queries.*
2. *If H is not bipartite, there exists c such that any 1-sided error tester requires at least $(\frac{c}{\epsilon})^{c\log(\frac{c}{\epsilon})}$ queries.*

Therefore, the bipartiteness of a graph can be characterized by the existence of 1-sided tester with query complexity polynomial in ϵ^{-1}.

In general, since a 1-sided error tester is immediately a 2-sided tester, it is not surprising that 2-sided testers may have lower query complexity. On the other hand, results on lower bounds for 1-sided testers are more common due to the more strict requirement.

4 Lower Bound on Triangle-Free Property Testing

In the next lecture, we will prove a lower bound on triangle-free property testing in the adjacency matrix model. Here we first introduce two tools that will be useful in the proof. The first tool is a theorem proved by Goldreich and Trevisan.

Theorem 7 *Let P be any graph property in the adjacency matrix model. Suppose T is a tester for P, with query complexity $q(n, \epsilon)$. Then there is a tester T' for P such that T' has the following form:*

1. *selects random sample of* $2q(n, \epsilon)$ *nodes,*

2. *queries all edges between the nodes,*

3. *makes Pass/Fail decision.*

Furthermore, if T *has 1-sided error, then so does* T'*.*

The theorem has the following implications:

- T' has query complexity $O(q^2(n, \epsilon))$,
- if T' needs $q'(n, \epsilon)$ queries, then T needs $\Omega(\sqrt{q'(n, \epsilon)})$ queries,
- T' is non-adaptive, that is, the queries it makes do not depend on previous query results.

Therefore, if we can prove that any tester for a certain graph property in the above special form has a query complexity lower bound in the adjacency matrix model, we can use that to derive a lower bound for any tester of the same graph property.

The second tool that we use is the Additive Number Theory Lemma.

Lemma 8 *For all* m*, there exists* $X \subseteq \{1, 2, \ldots, m\}$ *of size at least* $\frac{m}{e^{10\sqrt{\log m}}}$ *such that there is no solution in* X *to the equation* $x_1 + x_2 = 2x_3$ *with* $x_1 \neq x_2$*.*

We will continue the proof of lower bound in the next lecture.